Preprint
October 4, 2010

From “Oh, OK” to “Ah, yes” to “Aha!”: Hyper-systemizing and the rewards of insight

Chris Fields

Apdo 363-4013
Atenas 20501
Costa Rica

chrisfields38@gmail.com

Abstract

Hyper-systemizers are individuals displaying an unusually strong bias toward systemizing, i.e. toward explaining events and solving problems by appeal to mechanisms that do not involve intentions or agency. Hyper-systemizing in combination with deficit mentalizing ability typically presents clinically as an autistic spectrum disorder; however, the development of hyper-systemizing in combination with normal-range mentalizing ability is not well characterized. A review of anecdotal reports, survey-based measurements, and experimental studies of systemizing suggests the hypothesis that hyper-systemizing in the presence of normal-range mentalizing develops as an addiction syndrome driven by the positive affect associated with insight solutions. A neurocognitive model of hyper-systemizing as an outcome of insight addiction is constructed based on the incentive-sensitization model of addiction. If this model is correct, assaying subjects for bias toward systemizing or mentalizing would be expected to reveal significant activity differentials in temporal-parietal-frontal networks on cognitive tasks with systemizing or mentalizing components within the neurotypical population. Predictions of the model accessible to neurofunctional imaging, survey-based instruments and standard cognitive measures are outlined, and evidence pertaining to them considered.

Keywords: Systemizing, Mentalizing, Development, Addiction, Insight, Default network, Analogy

1. Introduction

The concept of “systemizing” was introduced by Baron-Cohen and colleagues to describe a problem-solving and explanatory strategy or style characterized by appeals to natural laws, physical mechanisms, algorithms or other concepts of causation that do not involve agency or intentions (Baron-Cohen, 2002; 2008). Systemizing or “mechanizing” (Crespi and Badcock, 2008) solutions and explanations are explicitly distinguished from “empathizing” or “mentalizing” solutions and explanations, which do appeal to causal agency and to actions taken to be guided by intentions, beliefs, desires, goals, fears, worries and other "folk psychological" attributes associated with agency by a theory of mind (ToM) system (Frith and Frith, 1999; 2003). Systemizing answers "how?" or "how did this come to be?" questions by tracing agency-independent causal mechanisms, without addressing why a particular mechanism is acting in any sense of "why?" that implies teleology or intention. As an experimentally-accessible construct capturing defining characteristics of scientific, technological, engineering and mathematical thinking (Baron-Cohen, 2008), the notion of systemizing drives a wedge between activities commonly thought of as explanation (Gopnik, 2000; Lombrozo, 2006), and suggests that humans may possess an integrated and specialized cognitive-affective "system brain" analogous to the "social brain" employed in mentalizing (Adolphs, 1999; Dunbar, 2003; Saxe *et al.*, 2004; Frith, 2007).

Experimental instruments have been developed to measure orientation or bias toward systemizing or mentalizing as a cognitive style (Baron-Cohen *et al.*, 2003); however, domain-independent experimental instruments to quantitatively measure systemizing or mentalizing ability or capability across broad populations are not available. Hence while it is clear that some individuals – e.g. theoretical physicists – are both highly oriented toward and highly capable of systemizing while others – e.g. novelists – are both highly oriented toward and highly capable of mentalizing, the relationships between bias and ability or capability within and between systemizing and mentalizing are not well understood. Experimental studies as well as common experience indicate that a bias toward systemizing typically, but not always, predicts a bias against mentalizing and vice-versa (Baron-Cohen *et al.*, 2002; 2003). Individuals with autism spectrum disorders (ASD) exhibit deficit mentalizing and a strong bias toward systemizing (Baron-Cohen *et al*., 2001; 2002; 2003); but the consequences of deficit systemizing, termed "system blindness" (Baron-Cohen *et al*., 2002), have not been investigated. Hence representations that combine bias towards and ability or capability in systemizing and mentalizing into a single cognitive dimension (e.g. Crespi and Badcock, 2008, Fig. 4) may be useful as heuristics, but both overstate and oversimplify current understanding of the positive and negative correlations between systemizing and mentalizing biases and capabilities.

Both males and females considered as subpopulations exhibit the full range of biases for or against systemizing; however, males as a group show a greater tendency toward systemizing than females as a group (Baron-Cohen *et al*., 2003; Goldenfeld *et al.*, 2006; Nettle, 2007). Differences in testosterone exposure *in utero* have been advanced as a possible explanation of this difference (Baron-Cohen *et al*., 2004). Consistent with a prenatal origin, the distinction between systemizing and mentalizing is observable early in development, initially in association with distinct orientations toward animate and inanimate objects (Karmaloff-Smith, 1995; Subrahmanyam *et al.*, 2002). A robust ability to attribute agency develops in infancy (Johnson *et al*., 2007; Saxe *et al*., 2005; 2007); infant tendencies toward mentalizing predict later childhood tendencies toward mentalizing (Wellman *et al*., 2004). An understanding that inanimate objects, unlike animate agents, respond in predictable ways to external causes develops in early childhood (Gopnik and Schultz, 2004). While infants separately categorize self-propelled inanimate objects and attribute internal casual powers to them (Subrahmanyam *et al.*, 2002; Luo *et al.*, 2009), an understanding that objects may exhibit spontaneous behaviors due to hidden, internal causes – arguably the earliest indication of robust systemizing – can typically be inferred around age four (Sobel *et al.,* 2007). A competitive mechanism for switching between systemizing and mentalizing strategies for problem solving is suggested by classic experiments with animated geometric figures, which demonstrate transitions between systemizing and mentalizing, in both children and adults, driven by minor changes in the motions of simple geometric shapes (Scholl and Tremoulet, 2000); such switching may be implemented by mirror-system neurons attuned to biological motion patterns (Puce & Perrett, 2003; Engel et *al*., 2007). A bias towards mentalizing correlates strongly with the Big Five personality factor Agreeableness; a bias toward systemizing correlates significantly with the Big Five factors Conscientiousness and Openness (Nettle, 2007), consistent with documented personality characteristics of scientists, and particularly of more creative scientists (Feist, 1998).

Individuals strongly biased toward systemizing are considered "hyper-systemizers" (Baron-Cohen *et al*., 2002; Baron-Cohen, 2008). While anecdotal evidence suggests a positive correlation between a strong bias toward systemizing and high-level capability in systemizing, this correlation has not been studied experimentally with large, unbiased populations; hence "hyper-systemizer" must be regarded as naming a cognitive orientation, not a highly-developed capability. The notion of a hyper-systemizer is, moreover, clearly culture- and even subculture-dependent: adopting a systemizing approach to human origins is unremarkable in contemporary western Europe, for example, but remarkable as a minority position in the United States (e.g. Paul, 2009); adopting a systemizing approach to human language understanding is commonplace in cognitive neuroscience or artificial intelligence circles, but is considered incoherent by many philosophers (e.g. Searle, 1980). Nonetheless, hyper-systemizers can be recognized by an unusually pervasive bias toward systemizing, e. g. as indicated by a Systemizing Quotient (SQ) score more than two standard deviations from the mean, typically combined with a bias against mentalizing (Baron-Cohen *et al*., 2002). The aversion to strong emotions, and to interpersonal conflict in particular among physicists noted by Feist and Gorman (1998) is consistent with hyper-systemizing. Hyper-systemizing in conjunction with deficit mentalizing ability presents in early development as ASD, with severity dependent on the severity of mentalizing deficit (Baron-Cohen et al., 2001; 2002; 2003; Crespi and Badcock, 2008; Ring *et al*., 2008); physical sciences, engineering and mathematics are common career choices of high-functioning individuals diagnosed with Asperger Syndrome or high-functioning autism (Baron-Cohen *et al*., 2001; Fitzgerald & O'Brien, 2007). It is not clear whether hyper-systemizing plays a causal role in the etiology of ASD, or is a developmental consequence of deficit mentalizing or other underlying deficits (Rejendran and Mitchell, 2006; Markram *et al*., 2007). Despite the strong correlation between hyper-systemizing and ASD (Baron-Cohen et al., 2001; 2002; 2003; Goldenfeld et al., 2006; Baron-Cohen, 2008), hyper-systemizing is also observed in both males and females with normal-range mentalizing ability (Baron-Cohen et al., 2003; Goldenfeld et al., 2006), as would be expected given the existence of highly systemizing-oriented scientists, engineers and mathematicians of both sexes who are able to attribute mental states to others, predict behavior on the basis of inferred intentions, and function effectively in large organizations. The mechanisms driving development of hyper-systemizing in conjunction with normal-range mentalizing ability, and in particular the motivational components of such asymmetric development, are not well characterized.

This paper suggests that the development of hyper-systemizing in conjunction with normal-range mentalizing is productively conceptualized as driven by addiction to the affective rewards experienced following insight. The term "addiction" here is not meant to indicate dysfunction, but rather to indicate a direct mechanism of activation of the reward pathway and consequent production of positive affect that is independent of natural rewards, and that is associated with strong attentional focus, high tolerance for risk, and over-learning of both endogenous and exogenous contextual cues predictive of reward (c.f. Hyman *et al*., 2006). In Section 2, differences between the affective experiences typically accompanying problem-solving by systemizing or mentalizing are reviewed. The affective spectrum from the "Oh, OK" of everyday solution-finding through the "Ah, yes" of mild surprise to the "Aha!" of insight is characterized in terms of increasing intensity of affective reward for current-state to goal-state conflict resolution. Historical and anecdotal evidence suggesting that hyper-systemizing individuals exhibit behavior patterns, including obsessive focus and risk tolerance, typical of addiction syndromes is then reviewed. In Section 3, the neurocognitive consequences that would be expected in an individual addicted to insight experiences are examined, primarily in the context of the incentive-sensitization model of addiction (Robinson and Berridge, 1993; 2008). It is shown that neurocognitive processes common to addiction syndromes would be expected to increase the salience of cues previously associated with systemizing, and specifically to decrease default-network activity. Default network activity is typically experienced as mentalizing; decreases in default-network activity correlate with both goal-directed, externally-focused problem solving and the formation of long-distance semantic links and structural analogies (Buckner et al., 2008; Kounios and Beeman, 2009). Hence systemizers in whom such an addictive process is active can be expected to become both more highly biased toward systemizing and more likely to form long-distance semantic links and structural analogies leading to successful insights. Section 4 outlines predictions of the addiction model of the development of hyper-systemizing that are accessible to neurofunctional imaging, survey instruments and standard cognitive tests. It suggests in particular that correlations with subject SQ scores may resolve the inter-subject variation observed in many functional imaging studies of problem solving into psychologically-meaningful evidence of neurocognitive diversity in the “neurotypical” population.

2. Affective experiences associated with systemizing

Common experience as well as developmental and adult cognitive studies suggest that many if not most people enjoy solving problems; it has been suggested that the human drive to explain is universal and analogous to the human drive for sex (Gopnik, 2000). The pleasure associated with discovery and understanding is regarded as an intrinsic motivation toward learning, not only in academic environments (Gottfried, 1985), but also in curiosity-driven unstructured play (Gibson, 1988; Karmaloff-Smith, 1995; Kaplan and Oudeyer, 2007). However, common experience also suggests that there are large individual differences in the extent to which the pleasure of learning is motivational, and desires that do not have pleasurable experience as primary components are often motivational in both academic (Covington, 2000) and general (Reiss, 2004a) environments. Consistent with such diversity, problem solutions or explanations do not uniformly induce pleasure; many religious explanations, for example, appear to be designed to induce fear, dread, or a social emotion such as affiliative solidarity (Reiss, 2004b). Hence, while a drive to explain may be universal, it appears to be a drive with multiple, separable components, and with multiple affective associations.

A diverse body of evidence indicates that feelings of pleasure are both more commonly and more likely to be associated with systemizing rather than mentalizing solutions or explanations. Mentalizing capacities develop in, and are generally regarded as having evolved in, small-group social contexts in which correct assessments of the intentions of others are critical for survival (Adolphs, 1999; 2003; Dunbar, 2003). Hence solutions to the most basic mentalizing problem – does this approaching person intend help or harm? - are naturally associated with primary emotions of affiliative bonding and fear. Solutions of more subtle mentalizing problems, such as determining whether a partner in a social exchange is cheating, are typically associated with social emotions such as, in this case, righteous anger or jealousy (Adolphs, 2003). Consistent with the strong emotional associations of mentalizing, expectation-reality conflicts in mentalizing often induce anxiety, and can induce pain comparable in character and intensity to physical pain (Eisenberger and Lieberman, 2004; Eisenberger, 2006). In contrast, the inanimate objects with which most systemizing is concerned tend not – with the exception of some products of technology – to be harmful or rewarding in and of themselves in the ways that animals and other humans are. Systemizing solutions and explanations can, therefore, be expected to lack the rich emotional tonality associated with mentalizing. Consistent with this expectation, the primary emotions associated with systemizing in self reports are pleasure and frustration (Shaw, 1999; Amabile *et al*., 2005). The dynamic range of these emotions is quite large: prominent historical figures from Archimedes onward have reported or displayed emotions ranging from mild pleasure to ecstatic elation following discoveries, and intense frustration bordering on despair when solutions seemed unreachable (Fitzgerald and O'Brien, 2007). Technical workplace subjects report feelings of pleasure ranging from “relieved and happy” to “all hyped” and “wonderful” accompanying successful systemizing, and varying levels of frustration accompanying failures (Amabile *et al*., 2005). Similarly, working scientists report frustration levels from “agitation” to “real bitter” in the face of seemingly intractable problems, and describe break-through insights as “really exciting” and “orgasm” (Shaw, 1999). The verbal richness of such reports indicate none of the alexithymia typically associated with ASD (Fitzgerald and Bellgrove, 2006); indeed, many subjects in these studies provide elaborate descriptions of “Aha!” contexts that clearly indicate competent mentalizing.

Personality characteristics and pathologies typical of hyper-systemizers and hyper-mentalizers (typically hypo-systemizers; Crespi and Badcock, 2008) provide supporting evidence that systemizing and mentalizing are associated with distinct affective spectra. Obsessive focus on work, social withdrawal, and neglect of personal maintenance are common to creative artists and scientists, but creative artists do not display the aversion to strong social emotions typical of creative scientists (Feist, 1998; 1999). Highly creative scientists report bouts of intense frustration, but clinical depression is more common in highly creative writers and artists (Feist, 1999; Nettle, 2001). Highly creative scientists may experience ecstatic pleasure following insights, but also report periods of uncertainty following the initial pleasure (Shaw, 1999) and demonstrate high levels of the Big Five factor Conscientiousness (Feist, 1998, Nettle, 2007). Highly creative artists demonstrate lower levels of Conscientiousness (Feist, 1998), and creative writers and artists are more prone to mania, and to psychotic spectrum disorders (PSD) in general (Nettle, 2001). It is interesting in this regard that while many prominent scientists have been retrospectively diagnosed with ASD (Fitzgerald & O'Brien, 2007), many prominent religious figures have been retrospectively diagnosed with epilepsy (Saver and Rabin, 1997), consistent with the more general association of hyper-religiosity with PSD (Nettle, 2001; Previc, 2006; Crespi and Badcock, 2008).

While the ability of putative hyper-systemizers to experience intense pleasure following insightful discovery has been much celebrated, a second characteristic of this population is less often noted: many take and have taken substantial risks, and many have died, in pursuit of explanations. Much scientific and technical work is inherently dangerous, and the dangers are often best understood by those who risk them. Marie Curie and Enrico Fermi, for example, are only the best known of the nuclear physicists who have died of radiation-associated cancers. Risky self-experimentation by naturalists and physicians has been commonplace throughout history. The pursuit of systemizing explanations has, moreover, been vigorously suppressed by governments and religious authorities until well into the 20th century, and the list of scientists and philosophers persecuted or executed for violating legal or socio-cultural sanctions against systemizing as a cognitive activity is long. It could argued, on a case-by-case basis, that social status, monetary benefit or other anticipated rewards motivated this history of risk taking, but the alternative hypothesis of intrinsic motivation is more consistent with cases for which self-reports are available (e. g. Fitzgerald and O'Brien, 2007). An understanding of hyper-systemizing as a cognitive-affective style must provide a natural explanation for this extraordinary tolerance of mortal risk.

Behavior patterns characterized by intense frustration and pleasure, obsessive focus on reward-producing activities in conjunction with social withdrawal, and a high tolerance for significant risk in the pursuit of rewards are suggestive of addiction as a mechanism (Hyman *et al*., 2006). The primary hypothesis of this paper is that the human drive to explain is implemented, in the case of systemizing but not of mentalizing, by the mesocorticolimbic reward pathway active in drug addictions: hyper-systemizers, on this hypothesis, are addicted to the experience of insight. The next section shows that this hypothesis not only explains the distinctive affective and personality associations of hyper-systemizing, but also provides mechanistic bases for the divergence of systemizing from mentalizing and for the characteristic, structural-analogy based creativity of hyper-systemizers.

3. Hyper-systemizing as addiction

Over the past decade, the previous primarily hedonic concept of addiction has been replaced by an integrated-systems concept that emphasizes reward-related learning, and in particular unconscious increases in salience of reward-related stimuli (Kelley and Berridge, 2002; Hyman, 2005; Nestler, 2005; Hyman *et al*., 2006; Grace *et al.*, 2007; Robinson and Berridge, 2008). The "incentive sensitization" model (Robinson and Berridge, 1993; 2008) has been instrumental in development of this new mechanistic concept of addiction. The primary components of the incentive sensitization model are that 1) a specialized mesocorticolimbic network encompassing nucleus accumbens (NAc), ventral tegmental area (VTA), amygdala, orbito-frontal cortex (OFC), dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC) and hippocampus mediates detection of, assignment of reward relevance and salience to, and learning and memory of cues associated with natural rewards including food and mating opportunities; 2) hedonic experience is generated in response to reward-related cues by activation of opiate and cannabinoid receptors in NAc and ventral pallidum, but is experienced in association with perceived cues and influences attentional focus via its representation in OFC and ACC (Berridge and Kringelbach, 2008; Kringelbach and Berridge, 2009; Smith *et al.*, 2009); 3) addictive drugs activate NAc and/or VTA directly, and induce both salience enhancement and associative over-learning of drug-related contextual cues and pleasurable drug-related experiences via induced activity of the pathway as a whole. Both the increased salience and pleasurable associations of drug-related cues contribute to the craving, obsession, and risk-tolerant reward-seeking behaviors typical of addictions.

While the incentive sensitization model was developed specifically to describe the affective and cognitive processes underlying addiction to drugs of abuse, the mechanisms that it describes can be expected to produce both over-learning of stimulus-related contextual cues and behaviors typical of addictions in response to any stimulus that activates the mesocorticolimbic reward pathway repeatedly with sufficient intensity. The core proposal of the model elaborated below is that in hyper-systemizers, task-completion signals generated by ACC in response to sudden, insightful problem solutions directly activate the mesocorticolimbic reward pathway with sufficient intensity to mimic the pathway-activating role of an addictive drug. In response to such activations, hyper-systemizers over-learn not just aspects of external context predictive of problem solving success, but also aspects of experienced cognitive context that are not only associated with, but that mechanistically facilitate, insightful problem solving.

The perceptual cues that contribute to representations of problems to be solved by mentalizing are in most cases, and arguably in all cases relevant to early development, associated with either natural rewards (e. g. smiling human faces) or natural dangers (e. g. angry human faces, snarling animals). In contrast, the perceptual cues that contribute to representations of problems to be solved by systemizing do not in general indicate natural rewards; in principle, *any* perceptual cue, however subtle, could contribute to the representation of a systemizing problem. Moreover, anticipated natural rewards associated with systemizing solutions, such as social status or economic rewards associated with successful technologies, may be distant in time, dependent on multiple external factors and hence highly uncertain, and in cultural contexts in which systemizing is frowned upon, fraught with danger. The pleasure associated with systemizing solutions is, therefore, difficult to explain in terms of immediate or even anticipated natural rewards. Such decoupling of pleasure from natural rewards suggests that an internal cognitive state generated in the course of problem solving itself, not a perceived or imagined natural reward, activates the reward system and hence generates the feeling of pleasure associated with successful systemizing. That an internally-generated problem-specific representation, not a percept or imagining, triggers the pleasure response associated with systemizing is further suggested by the fact that the sudden, intense pleasure of "Aha!" is associated not with the often-extended process of grappling with a problem, but rather with the recognition of a solution. Indeed, as discussed below, the pleasure response may be a component of solution recognition.

Multiple lines of evidence, all indirect, suggest that ACC is the locus of interaction between problem solving processes and the mesocorticolimbic reward pathway, and that an ACC-generated task-completion signal, in an appropriate context, is the internal representation that triggers an "Aha!" response. ACC monitors conflicts between cortical representations of current and goal states and evaluates the results of actions against the goals they are intended to achieve (Botvinick *et al*., 2004; Botvinick, 2007; Carter and van Veen, 2007), signaling both positive and negative progress toward a goal in a context-dependent way (Kennerley *et al*., 2006; Holroyd and Coles, 2008; Quilodran *et al.*, 2008). Reciprocal interactions between ACC and DLPFC implement attentional control (Ridderinkhof et al., 2004; Carter and van Veen, 2007); reciprocal interactions between ACC and OFC enable context-dependent re-evaluation of goals (Rushworth *et al*., 2007; Rolls and Grabenhorst, 2008). ACC task-monitoring signals to the basal forebrain modulate the intensity of attentional focus; feedback of task-monitoring signals to OFC via NAc and VTA enables the experience of attentional focus combined with expectation of success as concentration, task difficulty, motivation to continue, or fatigue (Sarter *et al.*, 2006; Grace et al., 2007; Boksem and Tops, 2008). Hence ACC measures progress, or lack thereof, toward a solution, and communicates this measurement to other components of the mesocorticolimbic reward pathway in a way that not only enables the use of such discrepancies to manage the problem-solving process, but also enables both the affective re-evaluation of goals and the affective experience of problem-solving progress or frustration.

Consistent with the role of ACC as an affect-mediating progress monitor, processes in which actual-state to goal-state conflicts are small are experienced as pleasurable and motivating, presumably via activation of dopamine-driven reward pathways (Kaplan and Oudeyer, 2007). Fluent performance of an effortful task requiring focused concentration can be intensely enjoyable, provided the performance remains fluent and conscious decision-making is not required (Csikszentmihalyi, 1996; Dietrich, 2004); the fact that the effortful activity itself, not just its successful outcome, is experienced as pleasurable indicates that in such cases process monitoring is directly coupled to affective experience. The specific dissociation of executive control and the feeling of effort in a patient with left ACC and OFC damage (Naccache *et al.*, 2005) confirms this direct link from ACC to the conscious representation of task progress. Positive affect correlates with ease of performance, and hence presumably with small actual-state to goal-state conflict signals from ACC, in tasks in which performance difficulty is varied without a subject's knowledge (Winkielman and Cacioppo, 2001; Winkielman *et al.*, 2003), and processing fluency may explain preferences for prototypes in cases ranging from facial features to abstract designs (Winkielman *et al*., 2006). These observations all suggest that positive affect increases as ACC-signaled discrepancy from a problem-solving goal decreases. If this is correct, the sudden convergence of a problem-solving process to its goal would be expected to induce a burst of positive affect, an "Aha!" experience.

If it is assumed that an ACC-generated task-completion signal indeed couples problem solving to the mesocorticolimbic reward pathway, the hypothesis that hyper-systemizing reflects addiction to insight becomes the hypothesis that, in at least some individuals with a pre-existing bias toward systemizing, ACC-generated task-completion signals can induce sufficiently strong responses in VTA and NAc to trigger enhanced positive valence assignment to and enhanced salience of cues associated with systemizing. On this model, routine solutions to routine problems would be expected to produce mild positive affect, but insightful solutions of hard problems, in particular, would be expected to provide sudden and strong task-completion responses and hence sudden and intense pleasure. One would expect, therefore, enhanced positive valence and enhanced salience of cues associated, in particular, with hard problems similar to ones previously solved with insight. Typical of drug addictions is the overvaluing of drug-related contextual cues, to the extent that they supplant even cues for natural rewards such as food and sex, with a resulting narrowing of focus of behavior (Hyman *et al.*, 2006). If insight addiction proceeds by a similar pathway, similar overvaluing of insight-problem related cues would be expected in this case. Hence by analogy with drug addictions, one would expect overvaluing and enhanced salience not only of specific problem-identifying cues, but also of contextual and enteroceptive cues associated with solving hard problems by insight.

Three aspects of context and enteroceptively-accessible cognitive "set" are consistent across anecdotal reports, survey-based measurements, and experimental studies of systemizing resulting in insight: moderate social withdrawal (Feist, 1998; Fitzgerald and O'Brien, 2007), positive affect (Shaw, 1999; Fredrickson, 2004; Amabile et al., 2005, Fitzgerald and O'Brien, 2007; Kounios and Beeman, 2009), and an attentional focus correlated with decreased default-network activation (Buckner *et al*., 2008; Kounios and Beeman, 2009). These effects are causally correlated. The default network links temporal-parietal junction (TPJ) areas implementing ToM to attention-control areas of medial prefrontal cortex including ACC (Raichle and Synder, 2007; Buckner *et al*., 2008). Default network activity is experienced as self-conscious and self-relevant reminiscence and future-oriented planning, typically in the modality of inner speech and accompanied by content-relevant emotions, mainly social emotions (Northoff *et al*., 2006; Raichle and Snyder, 2007; Buckner *et al*., 2008; Schilbach *et al.*, 2008). Default network activity is often obsessive, and high levels of default activity correlate with both major depression (Sheline et al., 2009) and schizophrenia (Kim et al., 2009). Artists often find inspiration in such emotionally-colored reflections (Nettle, 2001), but scientists tend toward discomfort with social emotions, particularly those associated with inter-personal conflict (Feist, 1998), and employ social withdrawal both to escape such emotions and to avoid distractions. Scientists are often happiest when "lost in their work," a flow-like state characterized by attentional focus on non-self-oriented elements of a task, and hence correlated with low default network activation (Buckner *et al*., 2008). It seems reasonable, therefore, to propose that decreased default network activity, experienced as a pleasant absence of uncomfortable and distracting self-oriented social emotions, is a highly-salient, over-learned cue associated with insight by hyper-systemizers.

Decreased default-network activity, and hence decreased mentalizing, self-referential thinking, and experienced social emotions (Buckner *et al*., 2008) appears to be not only a correlate of problem solving by insight, but an enabler and possibly a prerequisite of problem solving by insight. Experimental studies of insight have focused on the formation of distant semantic connections (Bowdon *et al*., 2005), a precursor of the formation of the structural analogies typical of insightful solutions of real-world problems (Gentner, 2003; 2005; Leech *et al.*, 2008). Formation of distant semantic connections requires activity in predominantly right-hemisphere association areas (Jung-Beeman *et al*., 2004; Kounios *et al.*, 2007; Sandkuhler and Bhattacharya, 2008) that overlap strongly with the default network (Bar, 2008). Recruiting these resources for systemizing would, therefore, require suppressing attention to default network activity. Interactions between ACC and rostral PFC appear to implement this suppression of default network activity in insight (Kounios *et al*., 2006; Subramaniam *et al.*, 2009), analogy (Green *et al*., 2006), and integration of solution components from different subtasks (De Pisapia and Braver, 2008). ACC-driven suppression of default network activity correlates with and may implement the facilitation of insight by positive affect (Subramaniam *et al.*, 2009; Kounios and Beeman, 2009). The involvement of rostral PFC in suppression of default network activity is consistent with its general role in attention switching and multi-tasking (Gilbert *et al.*, 2005; Dreher *et al*., 2008).

Decreased default-network activity also appears to be a general correlate of hyper-systemizing. High-functioning individuals with ASD tend to be hyper-systemizers (Baron-Cohen, 2002; Baron-Cohen *et al.*, 2002; 2003); default network activity is significantly lowered in ASD, as is the activity decrease accompanying default-network deactivation in attention-switching tasks (Kennedy *et al.*, 2006). Individuals with ASD exhibit low level of self-referential thinking (Lombardo *et al*., 2007) and functional differences in medial-frontal connection patterns that overlap the default network (Gilbert *et al*., 2009), both suggestive of reduced default-network activity as a correlate of (in this case pathological) hyper-systemizing. Hyper-systemizers without pathology appear more capable of maintaining a high degree of attentional focus on non-self-relevant stimuli (Billington *et al*., 2008), consistent with the obsessive attention to patterns typical of ASD (Baron-Cohen, 2002) and the obsessive attention to their work typical of creative scientists (Feist, 1998). Maintenance of externally-focused, non-self-relevant attention requires default-system deactivation (Buckner *et al.*, 2008). While these observations do not establish a direction of causation, they are consistent with decreased default-network activity and hence reduced mentalizing being a pre-requisite for successful systemizing and hence a learnable aspect of the experienced cognitive set associated with systemizing success.

The foregoing considerations all suggest that if the mesocorticolimbic addiction pathway is activated by ACC responses to sudden, insightful solutions to systemizing problems, the activity of this pathway can be expected to 1) increase any pre-existing bias toward systemizing; 2) decrease the frequency and intensity of mentalizing, self-referential thinking, and self-relevant emotional experience; and 3) increase the probability of recognizing distant semantic connections and analogies, all by the common mechanism of over-learned suppression of default-network activity. The hypothesis that hyper-systemizing is ACC-mediated addiction to insight thus provides a mechanism by which hyper-systemizing can develop in individuals with normal-range mentalizing ability, and explains both an increasing bias against mentalizing and increasing creativity and hence capability in solving systemizing problems in such individuals. Moreover, it explains common personality characteristics of hyper-systemizers, including their obsessive attention, tendencies toward social withdrawal and aversion to social conflict and social distractions. It is significant, moreover, that an addiction-driven mechanism would not be expected to develop hyper-mentalizing from a pre-existing mentalizing bias. As outlined above, mentalizing is associated with a broad spectrum of social emotions, many of them unpleasant; hence mentalizing cannot be expected to consistently activate the reward pathway.

4. Predictions of the addiction model of hyper-systemizing

The model of hyper-systemizing as an addiction developed above makes three general predictions. First, it predicts that hyper-systemizing is not exclusively an outcome of deficit mentalizing, as is observed in ASD, but can also develop in the presence of normal-range mentalizing. Second, it predicts that both the developmental progression and adult expression of hyper-systemizing are consistent with the typical phenomenology of addiction syndromes, and in particular display a primary dependence on endogenous rewards. While the model does not predict that learned associations with natural rewards are irrelevant, particularly in early development, it does predict that natural rewards play a progressively smaller role as hyper-systemizing develops from late childhood to early adulthood. Third, the model predicts that hyper-systemizing is implemented by two specific mechanisms: direct activation of the mesocorticolimbic reward pathway by ACC task-completion signals, and reward-driven over-learning of an experienced cognitive-affective context that includes suppression of default system activity. The latter mechanism, in particular, facilitates performance on systemizing tasks as discussed above; hence the model predicts that development of hyper-systemizing not only correlates with but causally enhances capability in systemizing.

The prediction that hyper-systemizing can develop in the presence of normal-range mentalizing distinguishes the addiction model from congenital-cause models, such as that of Crespi and Badcock (2008), that characterize hyper-systemizing as a developmental correlate or consequence of deficit mentalizing. The addiction model does not attempt to explain the origin of a bias toward systemizing in some individuals; such a bias may be innate, or may develop in response to differential experience with objects during infancy (Rakison and Yermolayeva, 2010). What the addiction model provides is a mechanism for amplifying a small initial bias toward systemizing into the strong bias characteristic of hyper-systemizing. The operation of this mechanism is independent of mentalizing capability; hence the addiction model predicts both that hyper-systemizers would display a distribution of mentalizing capabilities extending well into the normal range, and that the development of hyper-systemizing would not significantly negatively impact mentalizing capability, even if it significantly reduced the expression of mentalizing. The identification of hyper-systemizers within populations of “normal” subjects (Baron-Cohen *et al.*, 2002; 2003; Nettle, 2007) confirms the first of these predictions. Joint measurements of SQ and mentalizing capability, as opposed to mentalizing bias, at developmental timepoints from childhood to adulthood would test the second.

The prediction that hyper-systemizing develops by an endogenously-driven addiction mechanism distinguishes the current model from conventional learning models that emphasize learned associations with natural rewards over “intrinsic motivation” (e.g. Covington, 2000). The addiction model does not entail that natural rewards are irrelevant to the development of hyper-systemizing, but rather that natural rewards do not play a dominant role in the development of hyper-systemizing. Available evidence bearing on this question is primarily negative. One strand of evidence is provided by the intense and obsessive interests sometimes developed by preschool children. Intense childhood interests often involve functional objects with discernible parts and occur most frequently in males. Children who develop such interests typically do so suddenly and without unusually-high natural rewards from parents or others as a response; indeed children with intense preschool interests often abandon them due to negative reinforcement from teachers or peers (DeLoach *et al*., 2007). As is well documented in professional scientists, mathematicians and engineers, intense adulthood professional interests that involve systemizing are often accompanied by moderate to severe social withdrawal and neglect of social obligations and even self-maintenance (Feist, 1998; Baron-Cohen *et al*., 2002; Fitzgerald & O'Brien, 2007). Such behaviors are typical of addictions, are generally not rewarded socially, and are not rewarded monetarily with sufficient frequency or predictability to explain the behavior in terms of association with natural rewards. Association with natural rewards also does not readily explain the preference for challenging over easy problems often observed in individuals strongly oriented towards systemizing. Finally, association of systemizing as an activity with actual or anticipated natural rewards involves a form of self-relevant thinking, and hence would be expected to activate the default network (Buckner *et al*., 2008). The external attentional focus required for successful systemizing, however, specifically involves suppression of default-network activity (Buckner *et al.*, 2008; Kounios & Beeman, 2009). While these forms of evidence are indirect, all suggest that association with natural rewards cannot explain either the typical behaviors of hyper-systemizers or the neuro-functional characteristics of externally-directed attention or insight. Longitudinal studies of reward histories of hyper-systemizers, "balanced" individuals, and hyper-mentalizers from childhood to early adulthood would shed additional light on this issue, as would studies in which on-going work on both systemizing and mentalizing problems was interrupted by rewards or promises of rewards. The addiction model would predict that hyper-systemizers would display above average orientation to systemizing in childhood, but would be no more rewarded for systemizing, on average, than "balanced" individuals or hyper-mentalizers. It predicts that self-relevant representations of rewards, even if presented only as distractors, would disrupt performance on systemizing tasks more than on mentalizing tasks, and would disrupt the performance of hyper-systemizers more than that of other individuals.

The prediction that hyper-systemizing is implemented by ACC-driven reward signals and over-learned default network suppression distinguishes the current model from any alternative mechanistic model of the development of hyper-systemizing. Because default-network suppression is correlated with both externally-focused attention and enhanced ability to make long-distance semantic associations (Buckner *et al.*, 2008; Kounios & Beeman, 2009), the mechanism proposed by the addiction model entails that hyper-systemizing will correlate with enhanced systemizing ability and capability. Because the mechanisms proposed by the addiction model would not be expected to amplify a bias toward mentalizing into hyper-mentalizing, the addiction model also predicts that hyper-mentalizing will not, in general, correlate with enhanced mentalizing ability or capability. Hence the addiction model is inconsistent with unidimensional models such as that of Crespi and Badcock (2008) that propose a single mechanism to account for extremes of both systemizing and mechanizing.

Neurofunctional imaging and transcranial magnetic stimulation (TMS) based functional disruption methods have been used extensively to characterize the large-scale networks that implement the "social brain" and hence mentalizing as a problem solving strategy (Saxe *et al.*, 2004; Frith, 2007; Crespi & Badcock, 2008). Analogous studies employing systemizing problems would test the implementation-level predictions of the addiction model, specifically the prediction that ACC task-completion signals activate the reward network, and that systemizing is both enabled by and results in default network deactivation. Anatomical areas associated with the default network, particularly components of the temporal-parietal junction (TPJ), are known to overlap areas implicated in the ventral attention network (Decety and Lamm, 2007; Corbetta *et al*., 2008; Mitchell, 2008), as well as with association areas involved in planning, semantic association, and analogy (Bar, 2008). Mirror-neuron system components located in or near TPJ are activated during mentalizing (Saxe *et al.*, 2004; Frith, 2007; Crespi & Badcock, 2008), but are also known to be responsive to non-biological as well as biological motion patterns (Vingerhoets *et al.*, 2002; Schubotz and van Cramon, 2004; Engel *et al.*, 2007), to change their specificities in response to usage patterns (Catmur *et al.*, 2007; 2008), and to be involved in such typical systemizing tasks as estimating forces in physical interactions (Hegarty, 2004; Wolff, 2007) and solving mathematical problems (Qin *et al*., 2004; Cantlon et al., 2006). Adequately testing the prediction of default-network deactivation would, therefore, require designs that compared default network activity at high spatial resolution between systemizers and mentalizers, using both systemizing and mentalizing problem-solving tasks. The addiction model would predict activation of distinct components of TPJ during systemizing and mentalizing, with enhanced suppression of mentalizing-specific default-network components during performance of systemizing tasks in hyper-systemizers compared with "balanced" individuals or mentalizers. Daydreaming provides a passive alternative to problem solving for such studies. Daydreaming typically involves self-relevant thoughts and involves co-activation of the default network, ACC, and rostral PFC (Christoff *et al*., 2009); however, daydreaming studies to date have not separated subjects by systemizing or mentalizing bias. The addiction model would predict lower default-network activation and less self-relevant content in the daydreams of hyper-systemizers.

The addiction-model prediction that hyper-systemizing correlates with enhanced systemizing ability and hence problem-solving accuracy, but that hyper-mentalizing does not correlate with enhanced mentalizing ability, can be tested behaviorally by screening subjects in problem-solving ability studies for systemizing versus mentalizing bias. The recent observation that high systemizing bias correlates with high mental rotation ability (Cook and Saucier, 2010) is consistent with this prediction. Hyper-systemizers would be expected to perform better than "balanced" or mentalizer controls of matched verbal ability on semantic-association tests as employed in insight studies (Jung-Beeman *et al*., 2004; Kounios *et al*., 2007; Sandkuhler and Bhattacharya, 2008) and on standard verbal analogies (Gentner, 2003; Holyoak, 2005). Hyper-systemizers would also be expected to perform better than balanced or mentalizer controls of matched visuo-spatial ability on tool-improvisation problems, which are effectively visuo-spatial analogies (Fields, 2010). Conversely, hyper-mentalizers would expected to perform no better than balanced or systemizer controls on mentalizing tasks, such as correctly inferring the intentions of actors in stories.

5. Conclusion

A minority of the human population exhibits a strong bias toward systemizing, a problem-solving and explanatory style that relies on hypothesized physical mechanisms, not on intentional agency. The phenomenology of hyper-systemizing is suggestive of an addiction syndrome. The hypothesis that hyper-systemizing is driven by interactions between ACC and the mesocorticolimbic reward pathway active in drug addictions explains the obsessive attention and social withdrawal typical of hyper-systemizers, the increasing bias against mentalizing displayed by hyper-systemizers, and their tendency toward creativity, particularly in the generation of novel structural analogies. Successful systemizing requires suppression of the default network; hence the over-learned suppression of default network activity predicted by the addiction model provides a natural explanation for the positive correlation between a strong orientation toward systemizing and systemizing ability that is apparent among members of systemizing professions. The addiction model of hyper-systemizing makes a number of predictions amenable to experimental test. In particular, it predicts that presence or absence of natural rewards for systemizing will have little effect on hyper-systemizers, and that hyper-mentalizing will not correlate with high mentalizing ability.

The model of hyper-systemizing outlined here implies that hyper-systemizing is an instance of developmental specialization producing functionally-significant neurocognitive diversity. Spikins (2009; 2010) has suggested, based in part on the culturally-significant abilities associated with Asperger's Syndrome, that neurocognitive diversity may be a major driver of human evolution. For this to be the case, culturally-significant neurocognitive diversity must be both heritable and culturally tolerated. If the interpretation advanced here is correct, hyper-systemizing will appear in human populations provided small biases toward systemizing occur in the population, and these biases are amplifiable by the mesocorticolimbic reward pathway. In this case, the requirement for heritability is reduced to a requirement for a population-level distribution in systemizing bias, and the requirement for tolerance is reduced to a requirement for tolerance of small biases toward systemizing. Hyper-systemizing could appear, in such a population, without assortative mating between individuals biased toward systemizing, and even without cultural tolerance for hyper-systemizing. Similar mechanisms capable of amplifying small differences in network activity into culturally-significant neurocognitive diversity may be active in other areas as well.

Acknowledgements

Thanks to Spencer Dew and Eric Dietrich for discussions that contributed to the ideas developed in this paper, and to Penny Spikins for a preprint of Spikins (2010). Three anonymous referees provided suggestions that significantly clarified the presentation.

Conflict of Interest Statement

The author declares that he has no conflicts of interest relevant to the work presented in this paper.

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences 3*(12), 469-479.

Adolphs, R. (2003). Cognitive neuroscience of human social behavior. *Nature Reviews Neuroscience 4*, 165-178.

Amabile, T. M., Barsade, S. G., Mueller, J. S. & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly 50*, 367-403.

Bar, M. (2008). The proactive brain: Using analogies and associations to derive predictions. *Trends in Cognitive Sciences 11*(7), 280-289.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences 2*(2), 248-254.

Baron-Cohen, S. (2008). Autism, hypersystemizing, and truth. *The Quarterly Journal of Experimental Psychology 61*(1), 64-75.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders 31*, 5-17.

Baron-Cohen, S., Wheelwright, S., Lawson, J., Griffin, R. & Hill, J. (2002). The exact mind: Empathizing and systemizing in autism spectrum conditions. In U. Goswami (Ed.) *Handbook of Cognitive Development.* Oxford: Blackwell. (http://www.autismresearchcentre.com/docs/papers/2002_BCetal_Goswami.pdf)

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N. & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society of London B 358*, 361-374.

Baron-Cohen, S., Knickmeyer, R. C. & Belmonte, M. K. (2004) Sex differences in the brain: Implications for explaining autism. *Science 310*, 819-823.

Berridge, K. E. & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: reward in humans and animals. *Psychpharmacology 199*, 457-480.

Billington, J., Baron-Cohen, S. & Bor, D. (2008). Systemizing influences attentional processes during the Navon task: An fMRI study. *Neuropsychologia 46*, 511-520.

Boksem, M. A. S. & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews 59*, 125-139.

Botvinick, M., Cohen, J. & Carter, C. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences 8*(12), 539-546.

Botvinick, M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience 7*(4), 356-366.

Bowden, E. M., Jung-Beeman, M., Fleck, J. & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences 9*(7), 322-328.

Buckner R, Andrews-Hanna J & Schacter, D (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences 1124***,** 1-38.

Burgess, P. W., Simons, J., Dumontheil, I. & Gilbert, S. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. In J. Duncan, L. Phillips & P. McLeod (Eds.) *Measuring the Mind: Speed, Control, and Age* (pp. 217-248). Oxford University Press.

Cantlon, J., Brannon, E., Carter, E. & Pelphrey, K. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLOS Biology 4*(5) 0844-0854.

Carter, C. S. & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience 7*(4), 367-379.

Catmur, C., Walsh, V. & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current Biology 17*, 1527-1531.

Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M. & Heyes, C. (2008). Through the looking lass: Counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience 28*, 1208-1215.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences USA 106*(21), 8719-8724.

Cook, C. M. & Saucier, D. M. (2010). Mental rotation, targeting ability and Baron-Cohen's empathizing-systemizing theory of sex differences. *Personality and Individual Differences* (in press).

Corbetta, M., Patel, G. & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron 58*(3), 306-324.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology 51*, 171-200.

Crespi, B. & Badcock, C. (2008). Psychosis and autism as diametrical disorders of the social brain. *Behavioral and Brain Sciences 31*, 241-320.

Csikszentmihalyi, M. (1996). *Creativity.* New York: Harper Perennial.

De Pisapia, N. & Braver, T. S. (2008). Preparation for integration: The role of anterior prefrontal cortex in working memory. *NeuroReport 19*(1), 15-19.

Decety, J. & Lamm, C. (2007). The role of right temporoparietal junction in social cognition: How low-level computational processes contribute to meta-cognition. *The Neuroscientist 13*(6), 580-593.

DeLoach, J. S., Simcock, G. & Macari, S. (2007). Planes, trains, automobiles – and tea sets: extremely intense interests in very young children. *Developmental Psychology 43*(6), 1579-1586.

den Ouden, H. E. M., Frith, U., Frith, C. & Blakemore, S.-J. (2005). Thinking about intentions. *NeuroImage 28*, 787-796.

Dichter, G. & Belger, A. (2007). Social stimuli interfere with cognitive control in autism. *Neuroimage 35*(3) 1219-1230.

Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition 13*, 746-761.

Dreher, J.-C., Koechlin, E., Tierney, M. & Grafman, J. (2008). Damage to the fronto-polar cortex is associated with impaired multitasking. *PLOS One 3*(9), e3227.

Dunbar, R. (2003). The social brain: Mind, language, and society in evolutionary perspective. *Annual Reviews of Anthropology 32*, 163-181.

Eisenberger, N. I. & Lieberman, N. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Sciences 8*(7), 294-300.

Eisenberger, N. I. (2006). Identifying the neural correlates underlying social pain: Implications for developmental processes. *Human Development 49*, 273-293.

Engel, A., Burke, M., Fiehler, K., Bien, S. & Rosler, F. (2007). How moving objects become animated: The human mirror system assimilates non-biological movement patterns. *Social Neuroscience 3*, 368-387.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review 2*(4), 290-309.

Feist, G. J. (1999). Affect in artistic and scientific creativity. In S. W. Russ (Ed) *Affect, Creative Experience and Psychological Adjustment* (pp. 93-109). London: Taylor & Francis.

Feist, G. J. & Gorman, M. E. (1998). The psychology of science: Review and integration of a nascent discipline. *Review of General Psychology 2*(1), 3-47.

Fields, C. (2010). Implementation of structure mapping by event-file binding and action planning: A model of tool improvisation analogies. *Psychological Research* (in press) DOI=10.1007/s00426-010-0290-7.

Fitzgerald, M. & Bellgrove, M. A. (2006). The overlap between alexithymia and Asperger's Syndrome. *Journal of Autism and Developmental Disorders 36*(4) 573-576.

Fitzgerald, M. & O'Brien, B. (2007). *Genius Genes: How Asperger Talents Changed the World.* Shawnee Mission, KS: Autism Asperger Publishing.

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London B359*, 1367-1377.

Frith, C. (2007). The social brain? *Philosophical Transactions of the Royal Society of London B 362*, 671-678.

Frith, U. and Frith, C. (1999). Interacting minds – A biological perspective. *Science 286*, 1692-1695.

Frith, U. & Frith, C. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London B 358*, 459-473.

Gentner, D. (2003). Why we’re so smart. In: D. Gentner & S. Goldin-Meadow (Eds.) *Language and*

Mind: Advances in the Study of Language and Thought. Cambridge, MA: MIT Press. 195-235.

Gentner, D. (2005). The development of relational category knowledge. In: L. Gershkoff-Stowe & D. Rakison (Eds) *Building object categories in developmental time* (pp. 245-275). Hillsdale, NJ: Erlbaum.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology 39*, 3-41.

Gilbert, S., Frith, C. & Burgess, P. (2005). Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. *European Journal of Neuroscience 21*, 1423-1431.

Gilbert, S. J., Meuwese, J. D. I., Towgood, K. J., Frith, C. D. & Burgess, P. W. (2009). Abnormal functional specialization within medial prefrontal cortex in high-functioning autism: a multi-voxel similarity analysis. *Brain 132*, 869-878.

Goldenfeld, N., Baron-Cohen, S. & Wheelwright, S. (2006). Empathizing and systemizing in males, females, and autism. *Clinical Neuropsychiatry 2,* 338-345.

Gopnik, A. (2000) Explanation as orgasm and the drive for understanding: The evolution, function and phenomenology of the theory-formation system. In F. Kiel & R. Wilson (Eds.) *Cognition and Explanation* (pp. 299-323). Cambridge, MA: MIT Press.

Gopnik, A. & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences 8*(8) 371-377.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. Journal of Educational Psychology 77(6), 631-645.

Grace, A. A., Floresco, S. B., Goto, Y. & Lodge, D. J. (2007). Regulation of firing of dopaminergic neurons and control of goal-directed behavior. *Trends in Neurosciences 30*(5), 220-227.

Green, A., Fugelsang, J., Kraemer, D., Shamosh, N. & Dunbar, K. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research 1096,* 125-137.

Grimm, S., Ernst, J., Boesiger, P., Schuepbach, D., Hell, D., Boeker, H. & Northoff, G. (2009). Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. *Human Brain Mapping 30*(8), 2617-2627.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences 8* (6) 280-285.

Holroyd, C. B. & Coles, M. G. H. (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex 44,* 548-559.

Holyoak, K. (2005). Analogy. In: K. Holyoak and R. Morrison (Eds) *The Cambridge Handbook of*

Thinking and Reasoning. (pp. 117-142) Cambridge: Cambridge University Press.

Hyman, S. E. (2005). Addiction: A disease of learning and memory. *American Journal of Psychiatry 162*, 1414-1422.

Hyman, S. E., Malenka, R. C. & Nestler, E. J. (2006). Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience 29*, 565-598.

Johnson, S. C., Shimizu, Y. A. & Ok, S.-J. (2007). Actors and Actions: The role of agent behavior in infants' attributions of goals. *Cognitive Development 22*(3), 310-322.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arumbel-Liu, S., Greenblatt, R., Reber, P. J. & Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLOS Biology 2*(4), 0500-0510.

Kaplan, F. & Oudeyer, P.-Y. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience 1*(1), 225-236.

Karmaloff-Smith, A. (1995). *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: MIT Press.

Kelley, A. E. & Berridge, K. C. (2002). The neuroscience of natural rewards: Relevance to addictive drugs. *Journal of Neuroscience 22*(9), 3306-3311.

Kennedy, D., Redcay, E. & Courchesne, E. (2006). Failing to deactivate: Resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences USA 103*(21) 8275-8280.

Kennerley, S. W., Walton M. E., Behrens, T. E. J., Buckley M. J. & Rushworth M. F. S. (2006) . Optimal decision making and the anterior cingulate cortex. *Nature Neurosciences 9*(7), 940-947.

Kim, D. I., Manoach, D. S., Mathalon, D. H., Turner, J. A., Mannell, M., Brown, G. G., Ford, J. M., Gollub, R. L., White, T., Wible, C., Belger, A., Bockholt, H. J., Clark, V. P., Lauriello, J., O'Leary, D., Mueller, B. A., Lim, K. O., Andreasen, N., Potkin, S. G. & Calhoun, V. D. (2009). Dysregulation of working memory and default-mode networks in schizophrenia using independent component analysis, an fBIRN and MCIC study. *Human Brain Mapping 30*(11), 3795-3811.

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B. & Jung-Beeman, M. (2006). The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychological Science 17*, 882-890.

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M. & Jung-Beeman, M. (2007). The origins of insight in resting-state brain activity. *Neuropsychologia 46*, 281-291.

Kounios, J. & Beeman, M. (2009). The “Aha!” moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science 18*(4), 210-216.

Kringelbach, M. L. & Berridge, K. E. (2009). Toward a functional neuroanatomy of pleasure and happiness. *Trends in Cognitive Sciences 13*(11), 479-487.

Leech, R., Mareshal, D. & Cooper, R. (2008) Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences 31*, 357-378.

Lombardo, M., Barnes, J., Wheelwright, S. & Baron-Cohen, S. (2007). Self-referential cognition and empathy in autism. *PLOS One 9* (e883) 1-11.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences 10*(10), 464-470.

Luo, Y., Kaufman, L. & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. Cognitive Psychology 58, 441-486.

Markram, H., Rinaldi, T. & Markram, K. (2007). The Intense World Syndrome – an alternative hypothesis for autism. *Frontiers in Neuroscience 1*, 77-96.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex 18*, 262-271.

Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D. & Willer, J.-C. (2005). Effortless control: executive control and conscious feeling of mental effort are dissociable. *Neurophychologia 43*(9), 1318-1328.

Nestler, E. J. (2005). Is there a common molecular pathway for addiction? *Nature Neuroscience 8*(11), 1445- 1449.

Nettle, D. (2001). *Strong Imagination: Madness, Creativity and Human Nature.* New York: Oxford University Press.

Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology 98*, 237-255.

Northoff, G., Heinzel, A., de Greck, R., Bernpohl, F., Dobrowolny, H. & Panksepp, J. (2006). Self-referential processing in our brain – A meta-analysis of imaging studies on the self. *NeuroImage 31*, 440-457.

Paul G. (2009). The chronic dependence of popular religiosity upon dysfunctional social conditions. *Evolutionary Psychology 7*(3), 398-441.

Previc, F. H. (2006). The role of extrapersonal brain systems in religious activity. *Consciousness and Cognition 15*(3), 500-539.

Puce, A. & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Proceedings of the Royal Society of London B 358*, 435-445.

Qin, Y., Carter, C., Silk, E., Stenger, V. A., Fissell, K., Goode, A. & Anderson, J. R. (2004). The change in brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences USA 101*(15) 5686-5691.

Quilodran, R., Rothe, M. & Procyk, E. (2008). Behavioral shifts and action validation in the anterior cingulate cortex. *Neuron 57*, 314-325.

Raichle, M. E. & Snyder, A. Z. (2007). A default mode of brain function: A brief history of an evolving idea. *NeuroImage 37*, 1083-1099.

Rakison, D. H. & Yermolayeva, Y. (2010). Infant categorization. *Wiley Interdisciplinary Reviews: Cognitive Science 1* (in press).

Reiss, S. (2004a). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology 8*(3), 179-193.

Reiss, S. (2004b). The sixteen strivings for God. *Zygon 39*(2), 303-320.

Rejendran, G. & Mitchell, P. (2006). Cognitive theories of autism. *Developmental Review 27*, 224-260.

Ridderinkhof, K. R., van der Wildenberg, W. P. M., Segalowitz, S. J. & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition 56*, 129-140.

Ring, H., Woodbury-Smith, M., Watson, P., Wheelwright, S. & Baron-Cohen, S. (2008). Clinical heterogeneity among people with high-functioning autism spectrum conditions: Evidence favoring a continuous severity gradient. *Behavioral and Brain Functions 4*(11) PMC2265729.

Robinson, T. E. & Berridge, K. E. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews 18*, 247-291.

Robinson, T. E. & Berridge, K. E. (2008). The incentive sensitization theory of addiction: Some current issues. *Philosophical Transactions of the Royal Society B363*, 3137-3146.

Rolls, E. T. & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology 86*, 216-244.

Rushworth, M. F., Behrens, T. E., Rudebeck, P. H. & Walton, M. E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behavior. *Trends in Cognitive Science 11*, 168-176.

Sandkuhler, S. & Bhattacharya, J. (2008). Deconstructing insight: EEG correlates of insightful problem solving. *PLOS One 3*(1), e1459.

Sarter, M., Gehring, W. J. & Kozac, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews 51*, 145-160.

Saver, J. L. & Rabin, J. (1997). The neural substrates of religious experience. *Journal of Neuropsychiatry and Clinical Neurosciences 9*, 498-510.

Saxe, R., Carey, S. & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology 55*, 87-124.

Saxe, R., Tenenbaum, J. & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10- and 12-month-old infants. *Psychological Science 16*, 995-1001.

Saxe, R., Tzelnic, T. & Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental Psychology 43*, 149-158.

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R. & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and Cognition 17*, 457-467.

Scholl, B. & Tremoulet, P. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences 4* (8) 299-309.

Schubotz, R. & van Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observations and imagery. *The Journal of Neuroscience 24* (24) 5467-5474.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences 3*, 417-424.

Shaw, M. P. (1999). On the role of affect in scientific discovery. In S. W. Russ (Ed) *Affect, Creative Experience and Psychological Adjustment* (pp. 147-162). London: Taylor & Francis.

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., Mintun, M. A., Wang, S., Coalson, R. S. & Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceeding of the National Academy of Sciences USA 106*(6), 1942-1947.

Smith, K. S., Mahler, S. V., 2010Pecina, S. & Berridge, K. C. (2009). Hedonic hotspots: Generating sensory pleasure in the brain. In M. L. Kringelbach & K. E. Berridge (Eds.) *Pleasures of the Brain* (pp. 27-49). Oxford: Oxford University Press.

Sobel, D., Yoachim, C., Gopnik, A., Meltzoff, A. & Blumenthal, E. (2007). The blicket within: Preschooler's inferences about insides and causes. *Journal of Cognitive Development 8*(2) 159-182.

Somerville, L., Heatherton, T. & Kelley, W. (2006). Anterior cingulated cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience 9*, 1007-1008.

Spikins, P. (2009). Autism, the integrations of 'difference' and the origins of modern human behavior. *Cambridge Archaeological Journal 19*(2), 179-201.

Spikins, P. (2010). 'The world don't move to the beat of just one drum': Autism and the challenge of 'different minds' for the social sciences. Preprint, University of York.

Subrahmanyam, K., Gelman, R. & LaFosse, A. (2002). Animate and other separately-movable objects. In E. Fordes & G. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 341-373). London, England: Psychology Press.

Subramaniam, K., Kounios, J., Parrish, T. B. & Jung-Beeman, M. (2009). A brain mechanism for facilitation of insight by positive affect. *Journal of Cognitive Neuroscience 21*, 415-432.

Vingerhoets, G., de Lange, F., Vandemaele, P. Deblaere, K. & Achten, E. (2002) Motor imagery in mental rotation: An FMRI study. *NeuroImage 17*, 1623-1633.

Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S. & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science 7*(3), 283-288.

Winkielman, P. & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychphysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology 81*(6), 989-1000.

Winkielman, P., Schwarz, N., Fazendeiro, T. A. & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgement. In J. Musch & K. C. Klauer (Eds) *The Psychology of Evaluation: Affective Processes in Cognition and Emotion.* Mahwah, NJ: Lawrence Erlbaum.

Winkielman, P., Halberstadt, J., Fazendeiro, T. & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science 17*(9), 799-806.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology, General 136*, 82-111.